

EP0944245

Publication Title:

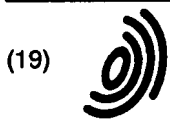
Hierarchical recursive motion estimator for video images encoder

Abstract:

By relaying on a temporal correlation among successive pictures beside on a spatial correlation of motion vectors of macroblocks of the currently processed picture and by the use of a hierarchical recursive motion estimation algorithm, the hardware complexity of video coders complying with the MPEG-2 standard can be greatly reduced without an appreciable loss of quality of the video images been transferred.

The method of motion estimation and a hardware embodiment of a coder are described and performance is compared with a prior motion estimation system.

Data supplied from the esp@cenet database - <http://ep.espacenet.com>



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 0 944 245 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
22.09.1999 Bulletin 1999/38

(51) Int. Cl.⁶: **H04N 5/00, H04N 7/36**

(21) Application number: **98830163.6**

(22) Date of filing: **20.03.1998**

(84) Designated Contracting States:
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE**
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant:
**STMicroelectronics S.r.l.
20041 Agrate Brianza (Milano) (IT)**

(72) Inventors:
• **Piccinelli, Emiliano
20052 Monza (IT)**

• **Pau, Danilo
20099 Sesto San Giovanni (IT)**
• **Zuccaro, Amedeo
28010 Fosseno di Nebbiono (IT)**

(74) Representative:
**Pellegrì, Alberto et al
c/o Società Italiana Brevetti S.p.A.
Piazza Repubblica, 5
21100 Varese (IT)**

(54) **Hierarchical recursive motion estimator for video images encoder**

(57) By relaying on a temporal correlation among successive pictures beside on a spatial correlation of motion vectors of macroblocks of the currently processed picture and by the use of a hierarchical recursive motion estimation algorithm, the hardware complexity of video coders complying with the MPEG-2 standard can be greatly reduced without an appreciable loss of qual-

ity of the video images been transferred.

The method of motion estimation and a hardware embodiment of a coder are described and performance is compared with a prior motion estimation system.

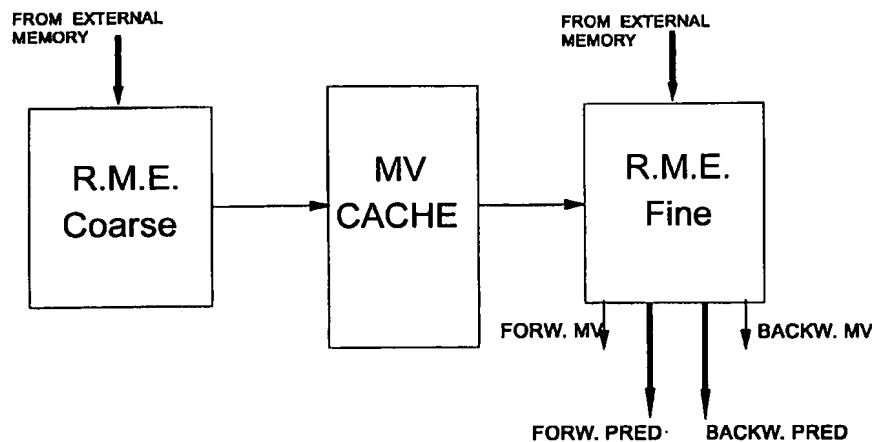


FIGURE 4

EP 0 944 245 A1

Description**FIELD OF THE INVENTION**

- 5 [0001] This invention relates to the algorithms and systems of motion estimation in processing video images and more in particular to an algorithm and an architecture of a motion estimator for implementing video coders compliant with the MPEG-2 standard.

DISCUSSION OF THE STATE OF THE ART ON MOTION ESTIMATION

10

[0002] The concept at the base of motion estimation is the following: a set of pixels of a field of a picture may be placed in a position of the subsequent picture obtained by translating the preceding one. Of course, these transpositions of objects may expose to the video camera parts that were not visible before as well as changes of their shape (e.g. zooming and the like).

- 15 [0003] The family of algorithms suitable to identify and associate these portions of images is generally referred to as "motion estimation". Such an association permits to calculate the portion of difference image by removing the redundant temporal information making more effective the subsequent process of compression by DCT, quantization and entropic coding.

[0004] Such a method finds in the standard MPEG-2 a typical example. A typical block diagram of a video MPEG-2 coder is depicted in Fig. 1.

20

[0005] Such a system is made up of the following functional blocks:

1) Field ordinator

- 25 [0006] This blocks is composed of one or several field memories outputting the fields in the coding order required by the MPEG standard. For example, if the input sequence is I B B P B B P etc., the output order will be I P B B P B B

- ♦ I (Intra coded picture) a field and/or a semifield containing temporal redundancy;
- 30 ♦ P (Predicted-picture) is a field and/or semifield from which the temporal redundancy in respect to the preceding I or P (previously co-decoded) has been removed;
- ♦ B (Birectionally predicted-picture) is a field and/or a semifield whose temporal redundancy in respect to the preceding I and subsequent P (or preceding P and successive P) has been removed (in both cases the I and P pictures must be considered as already co/decoded).

35

[0007] Each frame buffer in the format 4:2:0 occupies the following memory space:

40

45

50

standard PAL	720 x 576 x 8 for the luminance (Y) =	3,317,760 bit
	360 x 288 x 8 for the chrominance (U) =	829,440 bit
	360 x 288 x 8 for the chrominance (V) =	829,440 bit
	total Y + U + V =	4,976,640 bit
standard NTSC	720 x 480 x 8 for the luminance (Y) =	2,764,800 bit
	360 x 240 x 8 for the chrominance (U) =	691,200 bit
	360 x 240 x 8 for the chrominance (V) =	691,200 bit
	total Y + U + V =	4,147,200 bit

2) Motion Estimator

55

[0008] This is the block that removes the temporal redundancy from the P and B pictures.

3) DCT

[0009] This is the block that implements the cosine-discrete transform according to the MPEG-2 standard.

[0010] The I picture and the error pictures P and B are divided in 8*8 blocks of pixels Y, U, V on which the DCT transform is performed.

4) Quantizer Q

[0011] All 8*8 block resulting from the DCT transform is then divided by a quantizing matrix in order to reduce more or less drastically the magnitude of the DCT coefficients. In such a case, the information associated to the highest frequencies, less visible to human sight, tends to be removed. The result is reordered and sent to the successive block.

5) Variable Length Coding (VLC)

[0012] The codification words output from the quantizer tend to contain null coefficients in a more or less large number, followed by nonnull values. The null values preceding the first nonnull value are counted and the count figure constitutes the first portion of a codification word, the second portion of which represents the nonnull coefficient.

[0013] These paired values tend to assume values more probable than others. The most probable ones are coded with relatively short words (composed of 2, 3 or 4 bits) while the least probable are coded with longer words. Statistically, the number of output bits is less than in the case such methods are not implemented.

6) Multiplexer and Buffer

[0014] Data generated by the variable length coder, the quantizing matrices, the motion vectors and other syntactic elements are assembled for constructing the final syntax contemplated by the MPEG-2 standard. The resulting bit-stream is stored in a memory buffer, the limit size of which is defined by the MPEG-2 standard and cannot be overflowed. The quantizer block Q attends to the respect of such a limit, by making more or less drastic the division of the DCT 8*8 blocks depending on how far to the filling limit of such a memory buffer the system is and on the energy of the 8*8 source block taken upstream of the motion estimation and DCT transform process.

7) Inverse Variable Length Coding (I-VLC)

[0015] The variable length coding functions specified above are executed in an inverse order.

8) Inverse Quantization (IQ)

[0016] The words output by the I-VLC block are reordered in the 8*8 block structure, which is multiplied by the same quantizing matrix that was used for its precedent coding.

9) Inverse DCT (I-DCT)

[0017] The DCT transform function is inverted and applied to the 8*8 block output by the inverse quantization process. This permits to pass from the domain of spatial frequencies to the pixel domain.

10) Motion Compensation and Storage

[0018] At the output of the I-DCT block may alternatively be present:

- a decoded I picture (or semipicture) that must be stored in a respective memory buffer for removing the temporal redundancy in respect thereto from subsequent P and B pictures;
- a decoded prediction error picture (semipicture) P or B that must be summed to the information removed previously during the motion estimation phase. In case of a P picture, such a resulting sum, stored in dedicated memory buffer is used during the motion estimation process for the successive P pictures and B pictures.

[0019] These field memories are generally distinct from the field memories that are used for re-arranging the blocks.

11) Display Unit

[0020] This unit converts the pictures from the format 4:2:0 to the format 4:2:2 and generates the interlaced format for displaying the images.

[0021] The arrangement of the functional blocks depicted in Fig. 1, in an architecture implementing the above-described coder is shown in Fig. 2. A distinctive feature is in the fact that the field rearrangement block (1), the block (10) for storing the already reconstructed P and I pictures and the block (6) for storing the bitstream produced by the MPEG-2 coding, are integrated in memory devices external to the integrated circuit of the core of the coder, to which the decoder accesses through a single interface, suitably managed by an integrated controller.

[0022] Moreover, the preprocessing block converts the received images from the format 4:2:2 to the format 4:2:0 by filtering and subsampling of the chrominance. The post-processing block implements a reverse function during the decoding and displaying phase of the images.

[0023] The coding phase employs also the decoding for generating the reference pictures in order to make operative the motion estimation. For example, the first I picture is coded, thence decoded, stored (as described in paragraph 10)) and used for calculating the prediction error that will be used to code the subsequent P and B pictures.

[0024] The play-back phase of the data stream previously generated by the coding process uses only the inverse functional blocks (I-VLC, I-Q, I-DCT, etc.), never the direct functional blocks.

[0025] From this point of view, it may be said that the coding and the decoding implemented for the subsequent displaying of the images are nonconcurrent processes within the integrated architecture.

DESCRIPTION OF THE EXHAUSTIVE SEARCH MOTION ESTIMATOR

P field or semifield

[0026] Let us consider two fields of a picture (the same applies to the semifields), Q1 at the instant t and the subsequent field Q2 at the instant $t+(kp)*T$. kp is a constant dependant on the number of B fields existing between the preceding I and the subsequent P (or between two P), T is the field period (1/25 sec. for the PAL standard, 1/30 sec. for the NTSC standard). Q1 and Q2 are constituted by luminance and chrominance components. Let's suppose to apply the motion estimation only to the most energetic and therefore richer of information component, such as the luminance, representable as a matrix of N lines and M columns. Let us divide Q1 and Q2 in portions called macroblocks, each of R lines and S columns.

[0027] The results of the divisions N/R and M/S must be two integer numbers, not necessarily equal to each other.

[0028] Let Mb2(i,j) be a macroblock defined as the reference macroblock belonging to the field Q2 and whose first pixel, in the top left part thereof is at the intersection between the i-th line and the j-th column.

[0029] The pair (i,j) is characterized by the fact that i and j are integer multiples of R and S, respectively.

[0030] Fig. 2b shows how said reference macroblock is positioned on the Q2 picture, while the horizontal dash line arrows indicate the scanning order used for identifying the various macroblocks on Q2.

[0031] Let us suppose to project MB2(i,j) on the Q1 field, obtaining MB1(i,j).

[0032] Let us define on Q1 a search window having its center at (i,j) and composed of the macroblocks MBk(e,f) where k is the macroblock index. The k-th macroblock is identified by the coordinates (e,f), such that:

$$-p \leq (e-i) \leq +p \quad -q \leq (f-j) \leq +q$$

the indices e and f being integer numbers.

[0033] Each of said macroblock is said to be a predictor of MB2(i,j).

[0034] For example, if p=32 and q=48, the number of predictors is $(2p+1) * (2q+1) = 6,305$.

[0035] For each predictor, the norm L1 in respect to the reference macroblock is calculated; such a norm is equal to the sum of the absolute values of the differences between homologous pixels belonging to MB2(i,j) and to MBk(e,f). To each sum contribute R*S values, the result of which is called distortion.

[0036] Therefore $(2p+1) * (2q+1)$ values of distortion are obtained, among which the minimum value is chosen, thus identifying a prevailing position (e^*, f^*) .

[0037] The motion estimation process is not yet terminated because in the vicinity of the prevailing position, a grid of pixels is created for interpolating those that constitute Q1.

[0038] For example if Q1 is composed of:

.....

5 p31 p32 p33 p34 p35

 p41 p42 p43 p44 p45

10

[0039] After interpolation, the following is obtained:

15

20 p31 l1 p32

 l2 l3 l4

25 p41 l5 p42

30 where

$l1 = (p31 + p32)/2$

35 $l2 = (p31 + p41)/2$

$l3 = (p31 + p32 + p41 + p42)/4$

$l4 = (p32 + p42)/2$

40 $l5 = (p41 + p42)/2$

[0040] Let' us suppose to apply the above noted algorithm in the vicinity of the prevailing position by assuming, for example, $p=q=1$. In such a case, the number of predictors is equal to 8 and are constituted by pixels that are interpolated starting from the pixels of Q1. Let's identify the predictor with minimum distortion in respect to MB2(i,j).

[0041] The predictor more similar to MB2(i,j) is identified by the coordinates of the prevailing predictor through the above noted two steps of the algorithm.

[0042] The first step tests only whole positions while the second the sub-pixel positions. The vector constituted by the difference components between the position of the prevailing predictor and of MB2(i,j) is defined as the motion vector and describes how MB2(i,j) derives from a translation of a macroblock similar to it in the preceding field.

[0043] It should be noted that other measures may be used to establish whether two macroblocks are similar. For example, the sum of the quadratic values of the differences (norm L2). Moreover, the sub-pixel search window may be wider than what specified in the above example. All this further increases the complexity of the motion estimator.

[0044] In the example described above, the number of executed operations per pixel is equal to $6,305+8=6,313$, wherein each operation includes a difference between two pixels + an absolute value identification + a storage of the calculated result between the pair of preceding pixels of the same macroblock.

[0045] This means that to identify the optimum predictor there is a need for $6.313 \cdot S \cdot R$ parallel operators (at the pixel frequency of 13.5 MHz). By assuming $R=S=16$ as defined by the MPEG-2 standard, the number of operations required

is

$$6,313 \cdot 16 \cdot 16 = 1,616,128.$$

- 5 [0046] Each operator may function on a time division basis on pixels that belong to different predictors, therefore if each of these predictors worked at a frequency $4 \cdot 13.5 = 54$ MHz, the number of operators required would be $1,616,128/4 = 404,032$.

B field or semifield

10

[0047] Let us consider three picture fields (the same applies also to semifields), QP_{n-1} at the instant t , QBk_B at the instant $t+(k_B) \cdot T$ and QP_n at the instant $t+(k_P) \cdot T$ with k_P and k_B dependant on the number of B fields (or semifields) preventively selected. T is the field period (1/25 sec. for the PAL standard, 1/30 sec. for the NTSC standard). QP_{n-1} , QBk_B and QP_n are constituted by luminance and chrominance components. Let us suppose to apply the motion estimation only to the most energetic and therefore richer of information component, such as the luminance, representable as a matrix of N lines and M columns. Let's divide QP_{n-1} , QBk_B and QP_n in portions called macroblocks, each of R lines and S columns.

[0048] The results of the divisions N/R and M/S must be two integer numbers, not necessarily equal.

- 15 [0049] Let us $MB2(i,j)$ be a macroblock defined as the reference macroblock belonging to the field $Q2$ and whose first pixel, in the top left part thereof is at the intersection between the i -th line and the j -th-column.

[0050] The pair (i,j) is characterized by the fact that i and j are integer multiples of R and S , respectively.

[0051] Let's suppose to project $MB2(i,j)$ on the fQP_{n-1} field obtaining $MB1(i,j)$ and on the Qp_n obtaining $MB3(i,j)$.

- 20 [0052] Let's define on QP_{n-1} a search window with its center at (i,j) and composed of the macroblocks $MB1k(e,f)$ and on Qp_n a similar search window whose dimension may also be different, or in any case predefined, made up by $MB3k(e,f)$ where k is the macroblock index. The k -th macroblock on the QP_{n-1} is identified by the coordinates (e,f) , such that:

$$-p1 \leq (e-i) \leq +p1 \quad -q1 \leq (f-j) \leq +q1$$

- 30 while the k -th macroblock on the QP_n field is identified by the coordinates (e,f) such that:

$$-p3 \leq (e-i) \leq +p3 \quad -q3 \leq (f-j) \leq +q3$$

the indexes e and f being integer numbers.

- 35 [0053] Each of said macroblock is said to be a predictor of $MB2(i,j)$.

[0054] Thence, there are in this case two types of predictors for $MB2(i,j)$: those obtained on the field that temporally precedes the one containing the block to be estimated (I or P) referred to as "forward", and those obtained on the field that temporally follows the one containing the block to be estimated (I or P) referred to as "backward".

[0055] For example, if $p1=16$, $q=32$, $p2=8$, $q2=16$, the number of predictors is

40

$$(2p1+1) \cdot (2q1+1) + (2p2+1) \cdot (2q2+1) = 2,706.$$

[0056] For each predictor, the norm $L1$ in respect to the reference macroblock is calculated; such a norm is equal to the sum of the absolute values of the differences between homologous pixels belonging to $MB2(i,j)$ and to $MB1k(e,f)$ or $MB3k(e,f)$. To each sum contribute $R \cdot S$ values the result of which is called distortion.

45

[0057] Hence, we obtain the forward distortion values $(2p1+1) \cdot (2q1+1)$, among which the minimum value is chosen, thus identifying a prevailing position (e_F, f_F) on the field QP_{n-1} and $(2p2+1) \cdot (2q2+1)$ backward distortion values among which the minimum value is again selected, thus identifying a new prevailing position (e_B, f_B) on the QP_n field.

50

[0058] The motion estimation process is not yet attained because in the vicinity of the prevailing position, a grid of pixels is created to interpolate those that constitute QP_{n-1} and QP_n .

[0059] For example if QP_{n-1} is composed of:

55

.....

5 p31 p32 p33 p34 p35

 p41 p42 p43 p44 p45

10

15 [0060] After interpolation, we have:

.....

20 p31 l1 p32

 l2 l3 l4

25 p41 l5 p42

30 where

$$l1 = (p31 + p32)/2$$

$$35 \quad l2 = (p31 + p41)/2$$

$$l3 = (p31 + p32 + p41 + p42)/4$$

$$l4 = (p32 + p42)/2$$

$$40 \quad l5 = (p41 + p42)/2$$

[0061] Let us suppose to apply the above noted algorithm in the vicinity of the prevailing position by assuming, for example, $p=q=1$. In such a case, the number of predictors is equal to 8 and are constituted by pixels that are interpolated starting from the pixels of QP_{n-1} . Let's identify the predictor with minimum distortion in respect to $MB2(i,j)$.

45 [0062] In the same way we proceed for the QP_n field

[0063] The predictor more similar to $MB2(i,j)$ on QP_{n-1} and on QP_n is identified by the coordinates of the prevailing predictor through the above stated two steps of the algorithm predicted on each field. The first step tests only whole positions while the second the sub-pixel positions.

50 [0064] At this point we calculate the mean square errors of the two prevailing predictors (forward and backward), that is, the sums of the square of the differences pixel by pixel between the $MB2(i,j)$ with $(e_F \cdot, f_F \cdot)$ and with $(e_B \cdot, f_B \cdot)$.

[0065] Moreover, the mean square error between $MB2(i,j)$ is calculated with a theoretical macroblock obtained by linear interpolation of the two prevailing predictors and among the three values thus obtained we select the lowest.

[0066] Thence, $MB2(i,j)$ may be estimated using only $(e_F \cdot, f_F \cdot)$ or just $(e_B \cdot, f_B \cdot)$ or both though averaged.

55 [0067] The vector constituted by the components, difference between the position of the prevailing predictor and of $MB2(i,j)$, are defined as the motion vectors and describe how $MB2(i,j)$ derives from a translation of a macroblock similar to it in the preceding and/or successive field.

[0068] In the example described above, the number operations carried out for each pixel is equal to $2,706+8 \cdot 2=2,722$,

where each operation includes a difference between two pixels plus an absolute value plus an accumulation of the calculated result between the pair of preceding pixels and comprised in the same macroblock.

[0069] This means that for identifying the optimum predictor there is a need for $2,722 \cdot R \cdot S$ parallel operators (at the pixel frequency of 13.5 MHz). By assuming $R=S=16$ as defined by the MPEG-2 standard, the number of operations required is

$$2,722 \cdot 16 \cdot 16 = 696,832.$$

[0070] Each operator may function on a time division basis on pixels belonging to different predictors, therefore if each of them worked at a frequency of $4 \cdot 13.5 = 54$ MHz, the number of operators required would be $696,832/4 = 174,208$.

[0071] A high level block diagram of a known motion estimator based on an exhaustive search technique is depicted in Fig. 3, wherein the DEMUX block conveys the data coming from the field memory to the operators and where the MIN block operates on the whole of distortion values, calculating the minimum one.

PURPOSE AND SUMMARY OF THE INVENTION

[0072] The object of the present invention is to reduce the complexity of a motion estimator as used for example in an MPEG-2 video coder.

[0073] As an illustration of an efficient implementation of the method and architecture of the motion estimator of the present invention, a coder for the MPEG-2 standard will be taken into consideration. By the novel motion estimator of the invention is possible for example to employ only 6,5 operations per pixel in order to find the best predictor of the portion of picture currently being subjected to motion estimation, for an SP@ML compressed video sequence of either PAL or NTSC type. By contrast, the best result that may be obtained with a motion estimator of the prior art, would imply the execution of 569 operations per pixel beside the drawback of requiring a more complex architecture.

[0074] The method and relative architecture of motion estimation of the present invention, are defined in the appended claims.

[0075] The method of the invention implies a slight loss of quality of the reconstructed video images for the same compression ratio.

[0076] Nevertheless, such a degradation of the images is practically undetectable to human sight because the artifacts are distributed in regions of the images having a substantial motion content, the details of which practically pass unnoticed by the viewer.

DESCRIPTION OF A HIERARCHICAL RECURSIVE MOTION ESTIMATOR OF THE INVENTION

[0077] The number of operations per pixels required by the coding process may be remarkably reduced once the use of vectors calculated by the motion estimation process for macroblocks, spatially and temporally in the vicinity of the current macroblock is admitted.

[0078] The method herein disclosed is based on the correlation that exists among motion vectors associated to macroblocks in a homologous position in temporally adjacent images. Moreover, also the motion vectors associated to macroblocks belonging to the same picture, spatially adjacent to the current one may represent, with small errors, the motion of the current macroblock.

[0079] The process of motion estimation of the invention meets with the following requisites:

- ♦ The integration of the required number of operators necessary for implementing the method of motion estimation together with auxiliary structures, such as memories for allowing the reuse of precalculated vectors, must be markedly less onerous than that of motion estimators that do not avail themselves of the method of the invention;
- ♦ The loss of quality of the reconstructed images for a given compression ratio must be practically negligible, compared to motion estimators that do not implement the method of the invention.

[0080] In the ensuing description of the novel method of motion estimation, reference is made to a whole of fields equal in number to the distance (imposed beforehand and equal to M) between two subsequent P or I fields, included (a total number of fields equal to $M+2$ will be then taken into consideration according to the scheme of Figure 2b).

[0081] Let the temporal distance between two successive pictures be equal to a period of field. In particular let us suppose to have already considered the first, Q_{p-1} , a motion estimation in respect to the preceding (Q_0) and thereby consider its association to a motion field per macroblock. Said motion field is generated by using the same method of the first step, as described below.

FIRST STEP

[0082] Let us suppose to search on QP_{n-1} the prevailing macroblock predictor $MBQB(i,j)$ belonging to the QB1 field, that is the portion of Qp_{n-1} that more resembles it, and that such method is applied to all the QB1 macroblocks preceding it following a scanning order from left to right and from the top to the bottom.

[0083] According to Fig. 2c, let us consider:

$mv_MB5(i,j+S)$ is the motion vector associated to the macroblock belonging to QP_{n-1} and identified by the coordinates $(i, j+S)$

$mv_MB6(i+R, j)$ is the motion vector associated to the macroblock belonging to QP_{n-1} and identified by the coordinates $(i+R, j)$

$mv_MB3(i, j-S)$ is the motion vector associated to the macroblock belonging to QB1 and identified by the coordinates $(i, j-S)$

$mv_MB4(i-R, j)$ is the motion vector associated to the macroblock belonging to QB1 and identified by the coordinates $(i-R, j)$.

[0084] Let us consider, by way of example, to employ the above vectors for identifying, during a first phase, four predictors starting from the projection of $MBQB1$ on QP_{n-1} and that the prevailing predictor be identified by using the norm L1 (or the norm L2, etc.).

[0085] Generally, it is possible to use more than two predictors (belonging to QP_{n-1}) and also in a different number from those belonging to QB1. The above noted example has proven itself very effective during simulation.

[0086] The norm associated to the prevailing predictor is thereafter compared with precalculated thresholds derived from statistical considerations. Such thresholds identify three subsets, each composed of F pairs of vectors, wherein each pair, for example, is composed of vectors having components equal in terms of absolute value but opposite in sign. In the second step, such F pairs are summed to the vector that represents the prevailing predictor, identifying others $2 \cdot F$ predictors among which there may be also sub-pixels positions.

[0087] The prevailing predictor in the sense of the norm is the predictor of $MBQB1(i,j)$ on QP_{n-1} is the difference between their homologue coordinates individuate the motion vector to it associated.

[0088] It must be remarked the fact that in such a method, the norm is calculated starting from the result obtained by subsampling the macroblock according to a quincux scheme, or by interpolating the pixels of QP_{n-1} for generating predictor macroblocks disposed in sub-pixels positions.

[0089] The quincux grid is obtained by eliminating a pixel every two from the macroblock according to the following scheme:

<u>source macroblock</u>	<u>subsamped macroblock</u>		
A1 A2 A3 A4 A5 A6	A1	A3	A5
B1 B2 B3 B4 B5 B6	B2	B4	B6
C1 C2 C3 C4 C5 C6	C1	C3	C5

[0090] In this way, the operations necessary for calculating the norm are reduced by 50% compared to the case of an exhaustive search technique of a known motion estimator.

[0091] The method used for interpolating the pixels of QP_{n-1} , thus generating the sub-pixels thereof, is for example the one used in the exhaustive search estimator of the prior art.

[0092] Let's repeat what described above for QB1 also for the successive fields $QB2 \dots QB(M-1)$ and Qp_n , calculating the predictors of each on the respective fields immediately preceding temporally, obtaining so a motion estimator for each field of the partial sequence considered.

[0093] Said motion estimators must be stored in suitable structure to enable the second step.

SECOND STEP

[0094] In first place the QP_n field (type P) is coded and this requires to find a spreading of its macroblocks in respect to the Qp_{n-1} field positioned at a temporal distance equal to M field periods.

5 [0095] To perform this estimation let us consider the $MBP_n(i,j)$ block belonging to QP_n where i and j represent the position of the first top left pixel of the above mentioned macroblock in respect to the top left corner of the field it belongs to, and let us suppose that all the preceding QP_n macroblocks have already been submitted to such a process according to the scanning order.

10 [0096] By referring to Figure 2d, let us consider the two blocks of coordinates (i,j-S) immediately to the left and above (coordinates (i-R,j)) the block to be estimated $MBP_n(i,j)$ both belonging to QP_n and that have already been submitted to motion estimation to which they are therefore associated two motion vectors which will identify, on QP_{n-1} , two spatial predictors macroblocks.

[0097] Moreover, let us consider the field immediately preceding (in the temporal sense) the current one: $QB(M-1)$ has been already submitted to motion estimation in respect to its own previous field, $QB(M-2)$, and thus, to each of its
15 macroblock a translation vector is associated. A portion of such vectors may be considered to identify, properly scaled in terms of the temporal distance existing between QP_{n-1} and QP_n , the new $MBP_n(i,j)$ referred to as temporal predictors. These predictors are positioned on QP_{n-1} .

[0098] In particular, the positions identified by the motion vectors associated to the macroblocks, indicated in the figure with $T_{1,2}$ are tested if the temporal distance to estimate is of one field period. In this case only the vectors associated with T1 (of coordinates (i,j+S) and (i+R,j)) will be used, otherwise those indicated by T2 should be also considered
20 and whose coordinates are (i+R,j+2*S), (i+2*R,j+S), (i+2*R,j-S), (i+R,j-2*S).

[0099] Of course the number of these temporal predictors may also be different from the number indicated, however this choice is made based on the best experimental results.

[0100] Among all the indicated predictors only one is chosen using the criterion of the norm L1. This norm is then
25 compared with precalculated thresholds derived from statistical considerations. These thresholds identify 3 sub-sets of pair of vectors, whose components are equal in absolute value but with opposite signs. The number of such pairs is taken equal to F and F is the function of the temporal distance to cover by the estimation ($F=F(T_dist)$). In the second phase, such pairs are added to the vector that identifies the prevailing predictor, identifying other $2 \cdot F$ predictors among which there may be also subpixel positions.

30 [0101] The prevailing in the sense of the norm is the predictor of $MBP_n(i,j)$ on QP_{n-1} and the difference between their homologous coordinates identifies the motion vector to it associated.

[0102] For example, the number of operations per pixel according to the above described method for the P fields is equal to:

- first step	12
- second step	24
- coordinate position (i-0,j-0)	1
- partial total	37 (without quincux subsampling)
- final total	18.5 (with quincux subsampling)

45 [0103] This is followed by the estimation of the B fields: the procedure considers that the estimate is to be carried out both for the P or I field that temporally preceding the one to be estimated, in respect to both the I or P field that follows.

[0104] As for the estimation of the preceding I or p field the process is similar to what described above, whereas for the estimation towards the successive field (P or I) there are some differences in using the temporal predictors. In this
50 case, this term is used to identify the motion vectors associated to the macroblocks positioned in the same positions above described as for the temporal predictors of the P fields, though belonging to the immediately successive field (in the temporal sense) to the one to be estimated (thus always moving in the estimate direction).

[0105] For example, let us suppose we want to estimate $QB(M-2)$ in respect to QP_n . Hence, we utilize the vectors associated to the $QB(M-1)$ field. The latter are calculated during the implementation of the first algorithmic step.

55 [0106] It is necessary that such vectors are symmetrically overturned in respect to the origin because they identify the position of a block belonging to a future field as compared to a previous field. It is also necessary to scale them in a proper manner in function of the temporal distance between the current field and the one which to be estimated.

[0107] At this point, once the best backward predictor is chosen in the sense of the norm L1 between the two spatial

and temporal ones (for example 2 or 6), we choose again a certain number of pairs of small vectors symmetrical in respect to the origin (such a number is also a function of the temporal distance to cover) choosing them within the pre-defined whole by comparing the norm found with some thresholds defined by statistical considerations. Such pairs of vectors, added to the prevailing found above, will identify new predictors among which there may be also sub_pixel positions.

[0108] The prevailing in the sense of the norm is the final backward predictor for the block subject to estimation.

[0109] Finally, for each macroblock two predictors are so far identified, one on the I or P field that temporally precedes QB(M-2) and one on the successive I or P field. A third predictor is also identified and obtained by linear interpolation of the pixels belonging to the above cited predictors.

[0110] Out of the three predictors one is chosen based on the norm L1. The latter will be the final predictor which is subtracted from the reference block (the one submitted to estimation) obtaining in this way the prediction error.

[0111] For example, the number of operations per pixel according to the above described method is equal to:

- first step	12
- second step	33
- coordinate position (i-0,j-0)	1
- partial total	46 (without quincux subsampling)
- final total	23 (with quincux subsampling)

[0112] In these conditions, the performance in terms of signal/noise ratio that is obtained is equivalent to that of the known exhaustive search estimator (see Figure 3), while the complexity of the hardware implementation is markedly reduced.

BRIEF DESCRIPTION OF THE DRAWINGS

[0113] The different aspects and advantages of the invention will become even more evident through the following description of an embodiment and by referring to the attached drawings, wherein:

Figure 1 is a basic diagram of a video coder MPEG-2 MP@ML including the block called "motion estimator" which is the object of the present invention;

Figure 2 shows the architecture of the coder MPEG-2 MP@ML of Fig. 1;

Figure 2a is a reference scheme of the relative position of the macroblock taken into consideration in the description of the known method of motion estimation;

Figure 2b shows the temporal scheme of a whole of fields equal in number to a certain distance between subsequent P or I fields;

Figure 2c is a reference scheme of the relative position of the macroblock of pixels taken into consideration in the example of calculation according to the invention;

Figure 2d shows the relative position of the spatial and temporal macroblock predictors;

Figure 3 is a block diagram of the calculator of the norm between predictors and reference macroblocks, wherein is highlighted the array of parallel operator blocks that conduct the calculation of the norm L1;

Figure 4 shows the architecture of the hierarchical recursive motion according to the present invention;

Figure 5 shows the architecture of the estimator of Figure 4, relative to the first coding phase;

Figure 6 is a scheme of the quincux subsampler and interpolator;

Figure 7 shows the diagram of the block "comparator of M.A.E." for addressing the ROM contained in the diagram of Fig. 8;

Figure 8 is the diagram of the block "random addressing of macroblocks".

Figure 9 shows the memory architecture for the motion fields.

Figure 10 is the architectural scheme of the estimator of Figure 4, relative to the second coding phase;

Figure 11 shows the architecture of the estimator of Figure 4, relative to the implementation of the conclusive coding phase.

Figure 12 shows the block "random addressing of the macroblocks" of Figure 11.

ARCHITECTURE OF THE HIERARCHICAL RECURSIVE MOTION ESTIMATOR OF THE INVENTION

[0114] A block diagram of the hierarchical recursive motion estimator of the invention is depicted in Fig. 4.

[0115] In particular there are three blocks: the first carries out the first step of the procedure, that is the initialization and convergence of the motion fields; the third block carries out the second step of the algorithm, that is to say the coding itself of the MPEG-2 fields.

[0116] The above cited blocks interact through a memory that contains the two motion fields of the fields comprised between the first I or P field and a successive one of the same type.

[0117] The block referred to as **R.M.E. Coarse** is shown in Figure 5. This identifies a memory of $(N \cdot M)/(R \cdot S)$ cells (each of T bits) containing the motion vectors associated to the macroblocks preceding the current one and disposed on the same field and on the preceding one. Moreover, there is also a memory for storing the predictors belonging to the current field. This memory has dimensions $G \cdot H \cdot R \cdot S \cdot 8$ and permits to limit the number of accesses to the external memory, which would otherwise need to be accessed every time a predictor is required for feed the motion estimator, thus incrementing sensibly the passband.

[0118] By referring again to the same example described above, let us consider step 1 during which the four motion vectors are:

- mv_MB5(i, j+S)
- mv_MB6(i+R, j)
- mv_MB3(i, j-S)
- mv_MB4(i-R, j)

[0119] Depending on the position (i,j) of the macroblock which is being subjected to motion estimation and the reference macroblock, the motion vectors are acquired by the block **memory of the motion vectors** and are used for addressing the **macroblock memory**, from which said four macroblocks feed, one at a time, the **quincux subsampling block**. These subsampled macroblocks, eventually interpolated for defining the sub-pixel position, thereafter feed the block that calculated the norm L1 (or L2, etc.) between said predictor and the reference predictor. Said norm, by identifying the prevailing predictor of step 1 of the processing, permits to the **M.A.E. comparator** to address a ROM wherein vectors to be summed to the one associated to the prevailing predictor are stored.

[0120] The ROM is contained in the block called **random addressing of macroblocks**, at the output of which are obtained the addresses that are used for singling out the predictors in the **macroblocks memory**. These predictors feed the same blocks described in relation to the step 1. At the end of step 2, the motion vector V is obtained, and is stored in a register and made available to the coding process.

[0121] It may be noticed how the number of parallel operators is definitely very reduced, thus the implementation of the structure shown in Figure 3 is markedly less cumbersome simpler because the required number of operators is halved.

[0122] The structure of the block in Figure 4 called **MV cache** is shown in Figure 9: the motion vector outcoming the first block is conveyed to one of the six memories destined to contain the motion fields, each of which has a number of cells equal to $(N \cdot M)/(R \cdot S)$ of T bits each.

[0123] Such memories provide the motion fields used in the subsequent final estimation and in particular we have two output lines, one supplies the forward predictors and one supplies the temporally backward predictors.

[0124] The structure of the last block of Figure 4, called **R.M.E. fine**, is shown in Figure 10. It is possible to see how the motion vectors may be appropriately scaled in function of the estimate direction (forward or backward) and of the temporal distance and are then made available to the two forward and backward estimation blocks operating in parallel and whose structures are represented in Figure 11.

[0125] It can be noticed how the structure of the above cited estimation blocks is substantially similar to the one that operates the completion of the first estimation step described in Figure 5 with the exception of the absence of the memory dedicated to contain the motion field, being this contained in the **MV cache**.

[0126] Furthermore, the structure of the **block random addressing** is new, and its structure is shown in Figure 12. In the figure we notice that the adders exist in greater number as compared to the similar block existing in the structure of the first estimating step of Figure 8.

[0127] The adders serve to apply some small variations to the prevailing vector formed by testing of the spatial and temporal predictors. However, only a certain number of such adders is used. The selection is carried out based on the temporal distance to cover by the estimation (the greater is the distance, the greater is the number of adders used).

[0128] The selection of the type of variation (more or less important) is made by reading a ROM addressed by the MAE obtained from the **MAE comparator**. This ROM contains all the possible variations to be applied and is obtained through statistical considerations.

[0129] Figures 6 and 7 show the embodiments, respectively of the **quincux subsampler** and of the **MAE comparator** of the scheme of Figure 11, while the respective **calculator of the norm L1** has a functional scheme substantially identical to the one already shown in Figure 3.

[0130] With reference to the scheme of Figure 6, the **quincux subsampler** is formed by a plurality of 8-bit registers commanded, by way of a multiplexer (mux), by two signals of same frequency but opposite phase. The interpolator is constituted by a plurality of T registers, which permits the access to the sampled pixels at different instants, making them available for the downstream blocks of multiplication and addition. The coefficients C0, C1, C2, C3, C4 may for example take the following values, if applied to the source pixels p31, p32, p41, p42:

p31	p32	p41	p42	
1/2	1/2	0	0	----- I1
1/2	0	1/2	0	----- I2
0	0	1/2	1/2	----- I5
0	1/2	0	1/2	----- I4
1/4	1/4	1/4	1/4	----- I3
0	0	0	0	----- quincux subsampling implementation

[0131] The multiplexer finally selects the output, depending on the type of predictor required.

[0132] With reference to diagram of Figure 3, the circuit of calculation of the norm L1 among predictors and the reference macroblock is composed of a demultiplexer that conveys the predictors and the reference macroblock toward the appropriate operator. For example, if the macroblock has a 16*16 size, and by defining the norm L1 as the sum of the absolute values of the differences between homologous pixels (predictor/reference), the precision at the output of the subtractor block may be defined in 9 bits, in 8 bits the precision of the absolute value block and in 16 bits the precision of the accumulation block. The latter is constituted by an adder and a 16-bit register.

[0133] The outputs of the operators feed a block that calculates the minimum value, outputting the minimum value which is also called **MAE** (Mean Absolute Error).

[0134] With reference to the scheme of Figure 7 which shows the architecture of the **MAE comparator** for addressing the ROM of the scheme shown in Figure 9, the MAE must be comprised in one of the three subsets defined by the values:

$$- 0 + c_1$$

$$- c_1 + c_2$$

$$- c_2 + c_3$$

as a consequence at the output an address is produced.

[0135] By referring to Figure 8 showing the architecture of the "macroblocks random addressing", the address produced by the block of Figure 7 addresses a ROM which outputs 8 addresses, called "motion vectors", to be summed to the motion vector defined during step 1, as described above.

5 [0136] These sums are multiplexed for addressing the **macroblocks memory**.

[0137] By referring to Figure 9 that shows the memory architecture of the motion fields, a "demultiplexer" controlled by a counter, addresses the memory position where to store the single motion vector prevailing from the first step of the algorithm, writing thus the content of the single cache; as output, two multiplexers, both controlled by appropriate counters, selecting the vectors needed for the following estimation to be implemented in the second algorithmic step: 10 which are therefore required simultaneously, at the most two motion vectors: one for the "forward" estimation and one for the "backward" estimation.

[0138] By referring to Figure 10 showing the implementing architecture of the second coding step, the estimation direction, being it the forward or backward estimation and the temporal distance to cover modify respectively the sign and module, according to appropriate coefficients contained in the ROM, the temporal motion vectors read from the MV 15 cache: said vectors will be then used by the downstream structure which performs the final estimate phase which returns, as output, the prevailing vector and the motion vector associated to it.

[0139] Finally, it should be noticed that the two predictors (forward and backward) are added so to generate the "interpolated" predictor.

[0140] Figure 11 shows the architecture of the final phase block of the motion estimation which is similar to that of the analogous block of Figure 5 relative to the implementation of the first algorithmic step with the exception of the macroblocks random addressing block, which has a further input: T_dist. This is required for the selection of the total number of variations to be applied to the prevailing vector, following to the spatial/temporal predictors test.

[0141] Figure 12 shows the macroblock random addressing macroblock wherein to the prevailing motion vector are added the variations selected on statistical factors, direction of the estimate and temporal distance to cover.

25 [0142] The embodiments and applications of the motion estimator of the invention are numerous, among these the following can be mentioned:

- The motion estimation may be implemented by extracting predictors from a temporally preceding picture and also from a temporally successive picture. If both estimations are implemented in parallel, replicas of the structure of 30 Figure 4, operating in parallel, may be employed. The use of replicas of the motion vector memory and of the macroblocks memory is also contemplated.
- Coders for recording on digital video disks also called DVD RAM.
- 35 • Camcorders.
- Digital coders, even if not based on the MPEG-2 standard, where a step of motion estimation is required.

Claims

- 40 1. A method of motion estimation from homologous fields or pictures of successive images for video coders, comprising dividing each picture of N lines and M columns to be subjected to motion estimation in a plurality $(N/R) \cdot (M/S)$ of reference macroblocks, each of R lines and S columns of pixels, defining for each of said reference macroblocks a set of $P+Q+2 \cdot D$ predictor macroblocks of at least one of luminance and chrominance components disposed on 45 the preceding picture, and wherein:

a) P predictors are identified using motion vectors associated to macroblocks that precede the reference macroblocks on the same picture according to a raster scanning, by projecting said vectors on the macroblock homologous to the reference macroblock and disposed on the preceding picture,

50 b) Q predictors are identified by the use of motion vectors associated to macroblocks following the reference macroblock on the preceding picture according to a raster scanning, by projecting said vectors on the macroblock homologous to the reference macroblock placed on the preceding picture,

55 c) $2 \cdot D$ predictors are singled by the use of the motion vectors associated to P and Q predictor macroblocks summed to integer and/or fractionary quantities predefined in look-up tables addressed by an integer number associated to the field between a number C of available fields of the norm L1 with minimum P and Q values and the reference macroblock,

d) a subsampling block is sequentially fed with the predictor values of said $P+Q+2^*D$ macroblocks and with the pixel values of the reference macroblock,

e) for each pair of macroblocks constituted by one and only one among the predictors P or Q or 2^*D and the reference macroblock, the norm referred to as L1 is calculated and represents the sum of the differences in absolute value between the homologous pixels belonging to said pair of macroblocks,

f) the minimum value of said norms is identified and the x and y components are calculated for the motion vector associated to the reference macroblock as the difference between the homologous coordinates of the first pixel, according to a raster scanning of the reference macroblock, and the predictor with minimum distortion which minimizes the value of said norm L1,

wherein the total number of operations is reduced by employing a recursive procedure based on the correlation existing among motion vectors associated to macroblocks adjacent to the reference macroblock during a picture processing, storing the motion vectors associated to all the macroblocks of picture, overwriting them one by one with the motion vectors associated to corresponding macroblocks of a successive picture during the processing thereof and as said motion vectors are calculated in succession following a raster type of scanning of the macroblocks and using the stored motion vectors already calculated, associated to said P predictors macroblocks and the stored motion vectors associated to said Q macroblocks, nonhomologous to said P macroblocks adjacent to the reference macroblock during the processing of the current picture for addressing $P+Q$ predictor values, characterized in that it also comprises the following operations:

generating two motion vectors (V) for each macroblock of the field being processed, storable in two distinct output registers, through two parallel recursive search estimators, each estimator being fed through multipliers with motion vectors read from a B number of memories containing $(N/R)*(M/S)$ motion vectors;

said motion vectors supplied to said two estimators being multiplied by precalculated constants (T_dist) contained in a read only memory whose value depends on which of said B memories the motion vectors are read, from the I, P or B field being estimated, and on whether the estimation is carried out in relation to a preceding or successive field to the one currently under estimation;

said B memories being periodically overwritten with the motion vector values calculated in succession for the macroblocks relative to the future fields to the field being estimated, correspondingly to the enabling period of the B memories and according to the coding process specified in a), b), c), d), e) and f).

2. The method of claim 1, further comprising the following steps:

identifying the predictor macroblock with minimum distortion among all the above noted $P+Q$ predictor value;

comparing the so-called L1 norm value associated to said minimum distortion predictor with a plurality of T precalculated thresholds, derived from statistical considerations, identifying a plurality D of pairs of vectors, each pair constituted by vectors having components of identical absolute value but of opposite sign;

summing said pairs of vectors to the vector associated to the said minimum distortion predictor, identifying a number 2^*D of predictors, double in respect to the number of said pairs of vectors, by including intermediate or sub-pixel positions;

calculating said L1 norm value for each pair of macroblocks constituted by one predictor macroblock belonging to said 2^*D set and said reference macroblock;

identifying the minimum distortion macroblock among all the 2^*D macroblocks;

calculating the motion vector.

3. The method of any one of the preceding claims, characterized in that said L1 norm value is calculated from the result obtained by subsampling the macroblock according to a quincux grid.

4. The method according to any of the claims 1 and 2, wherein the L1 norm value is calculated by interpolating the pixels of the predictor by generating a new predictor in a sub-pixel position in respect to the plurality of positions

associated to the preceding picture.

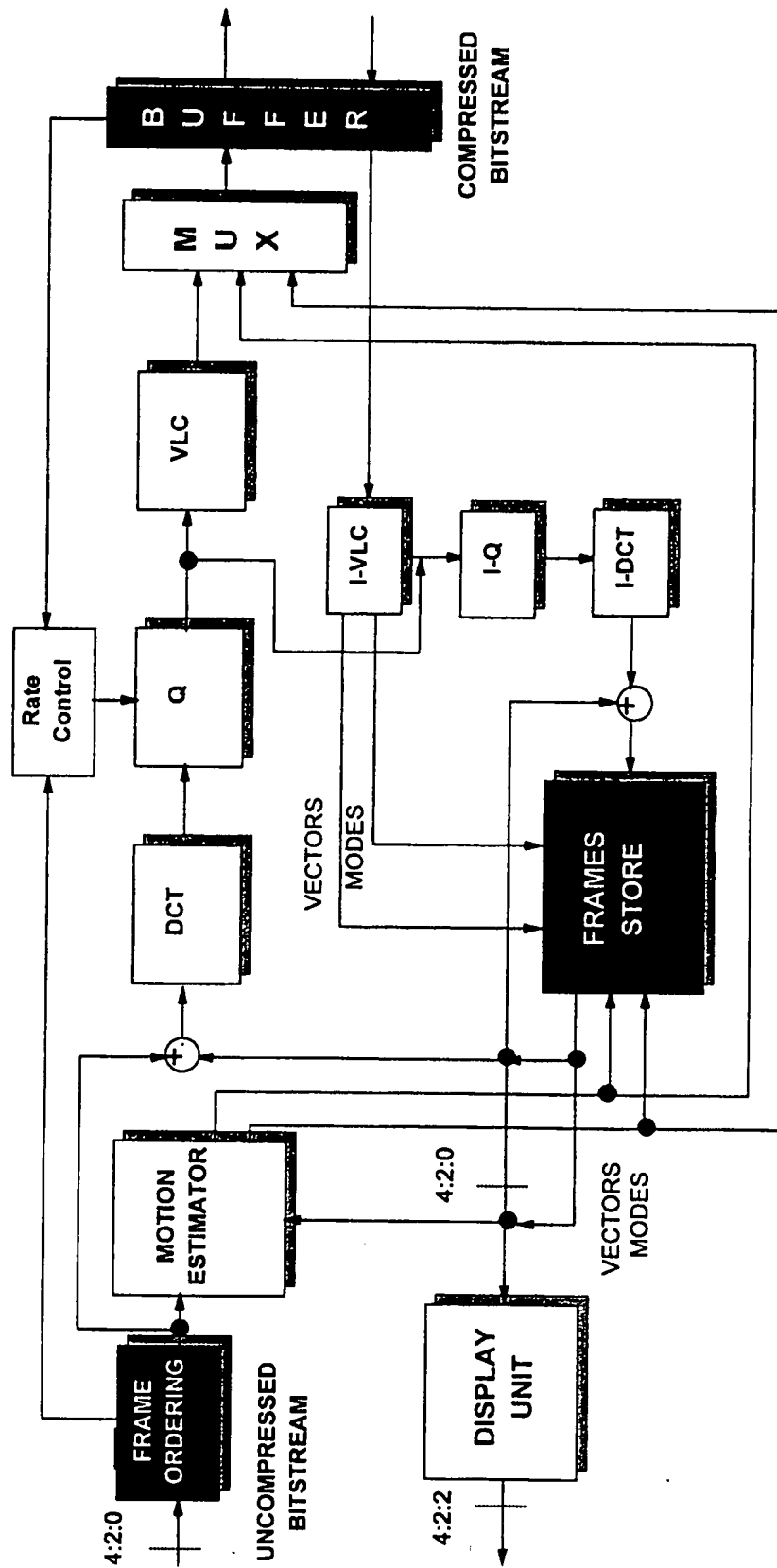
5. A recursive motion estimator for generating two distinct motion vector (V) storable in respective output registers, characterized by comprising:

a first recursive estimation block (R.M.E. coarse) generating a motion vector for each macroblock belonging to the field currently estimated and storing the vectors in a B number of memories belonging to a memory vector cache (MV CACHE) and activated in succession at every B field period;

said memory vector cache (MV CACHE) feeding a second recursive estimation block (R.M.E. fine) with at least two motion vectors contained in one of said B memories currently activated, said second recursive estimation block (R.M.E. fine) comprising:

a multiplier circuit of the two vectors fed to said recursive estimation block (R.M.E.), respectively by two coefficients (T_dist1, T_dist2) read from a read only memory whose value depends on the type of field being estimated, on the field that contains the predictor, on the respective temporal distance and on the fact of whether the predictor field follows (T_dist1) or precedes (T-dist2) the one being estimated;

a pair of recursive search estimators, the first operating on future predictor fields and the second on past predictor fields, input with the respective vectors resulting from said multiplication and with the respective coefficients (T_dist1, T_dist2)



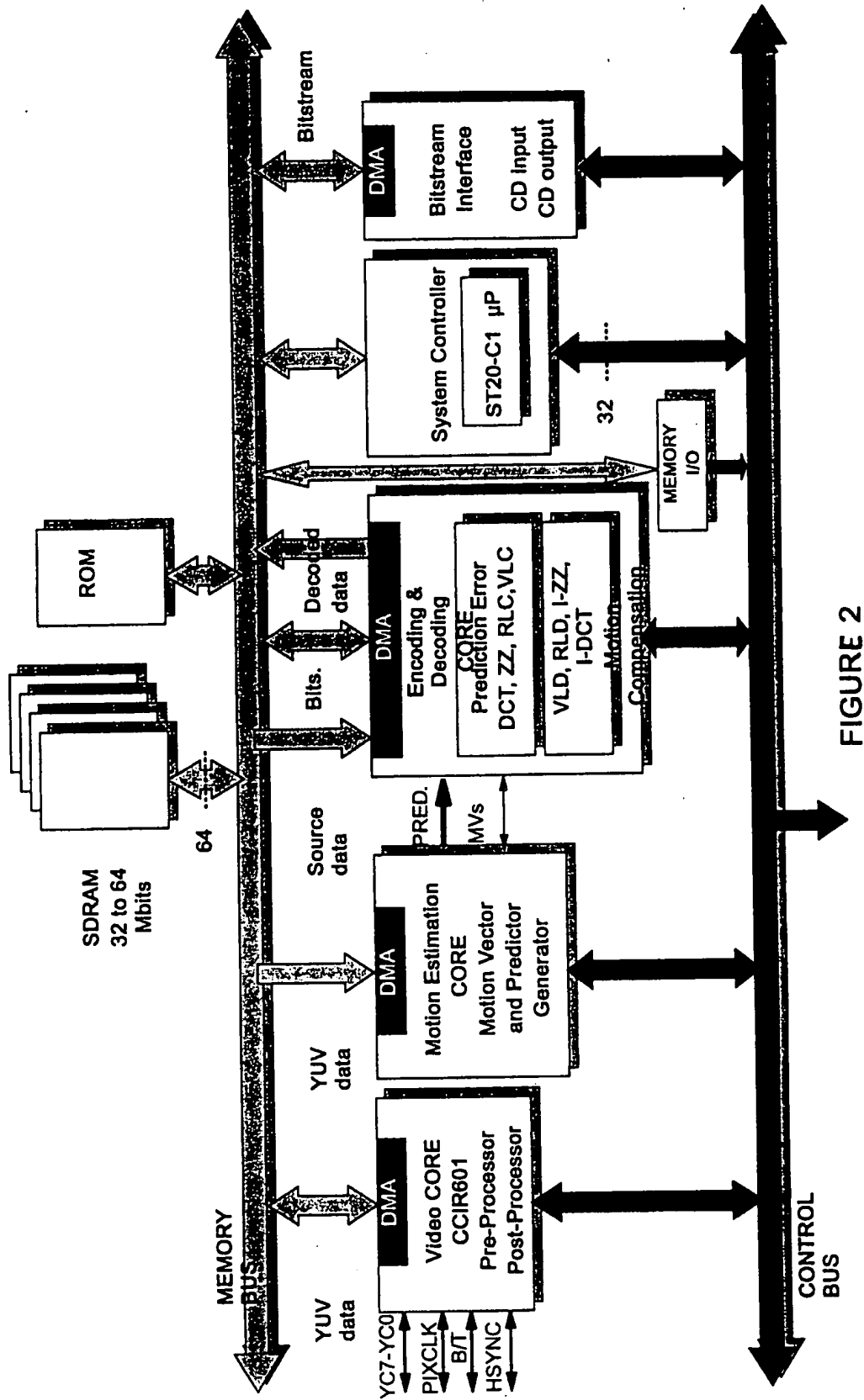


FIGURE 2

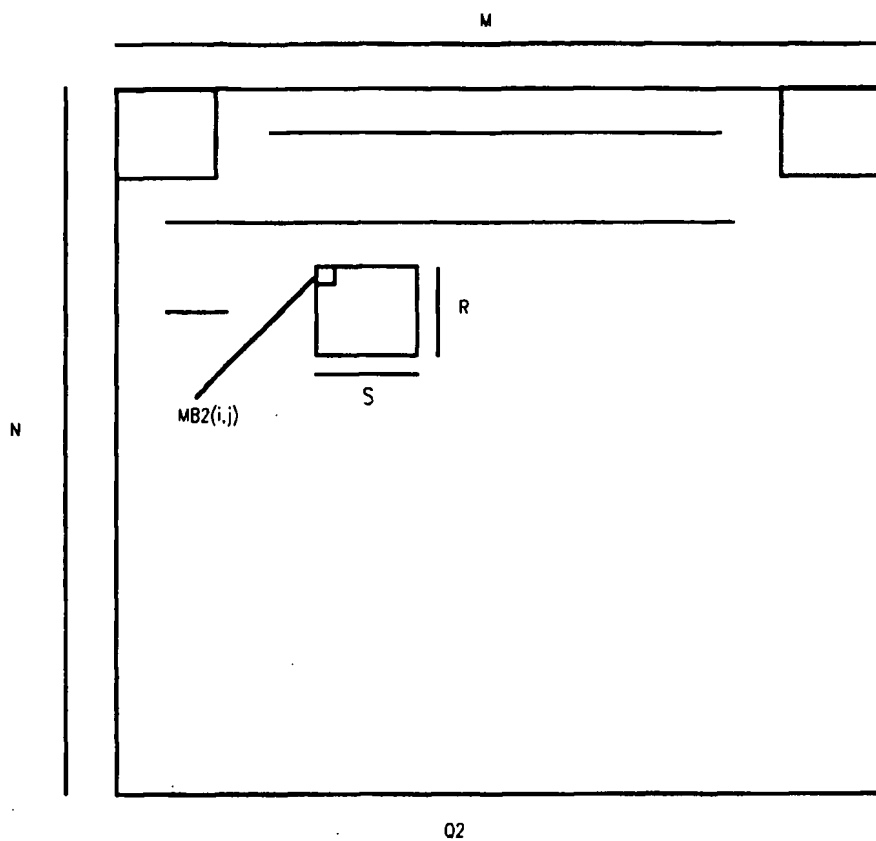


FIGURE 2a

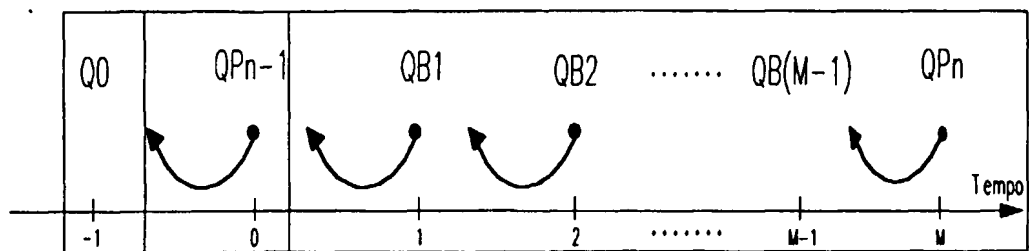


FIGURE 2b

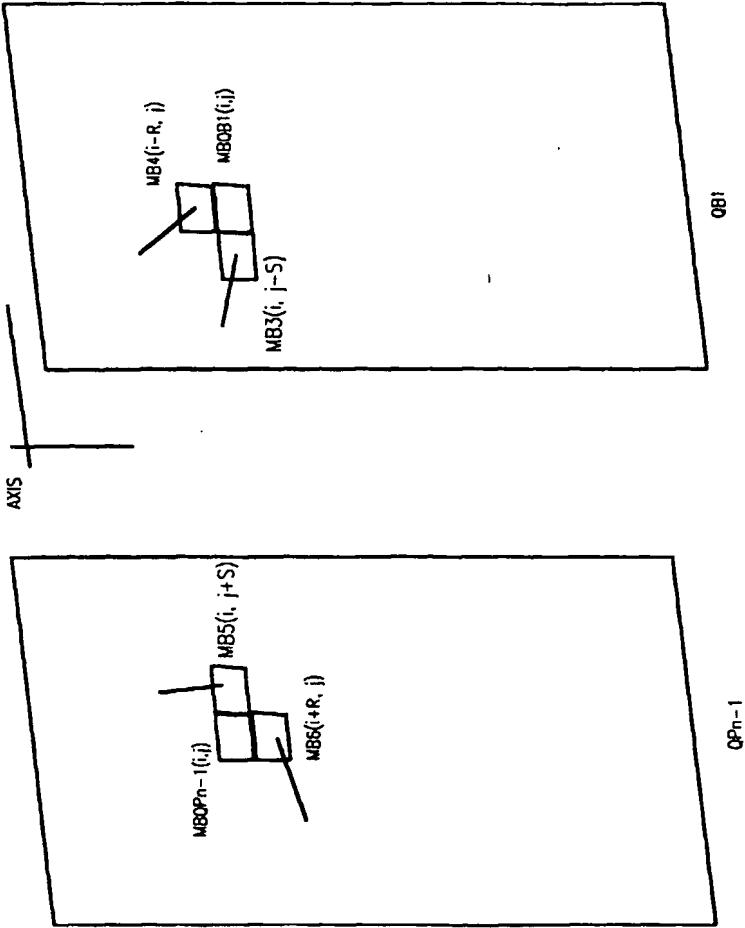


FIGURE 2c

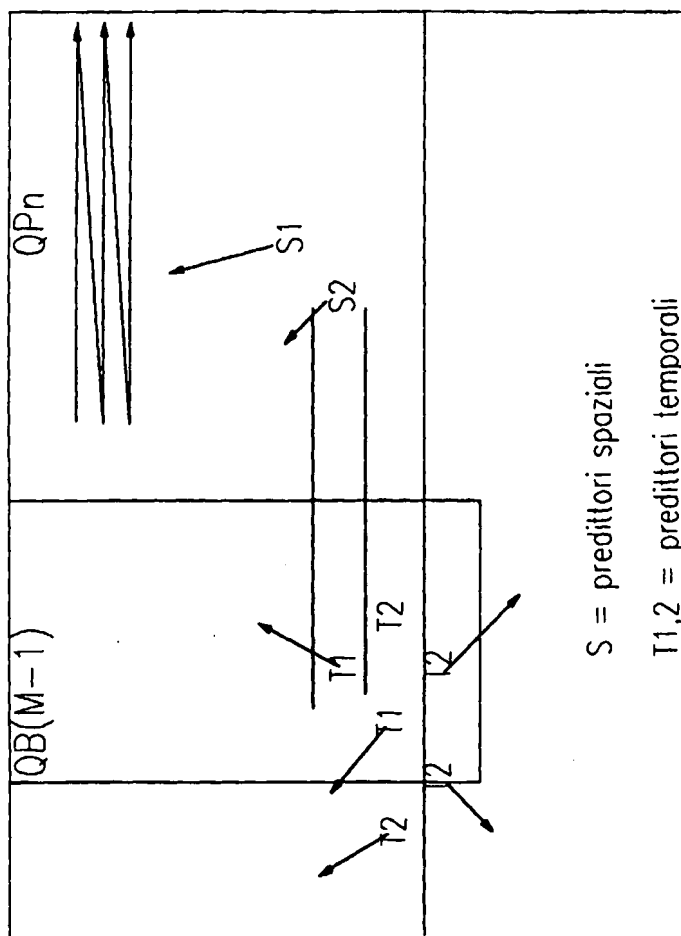


FIGURE 2d

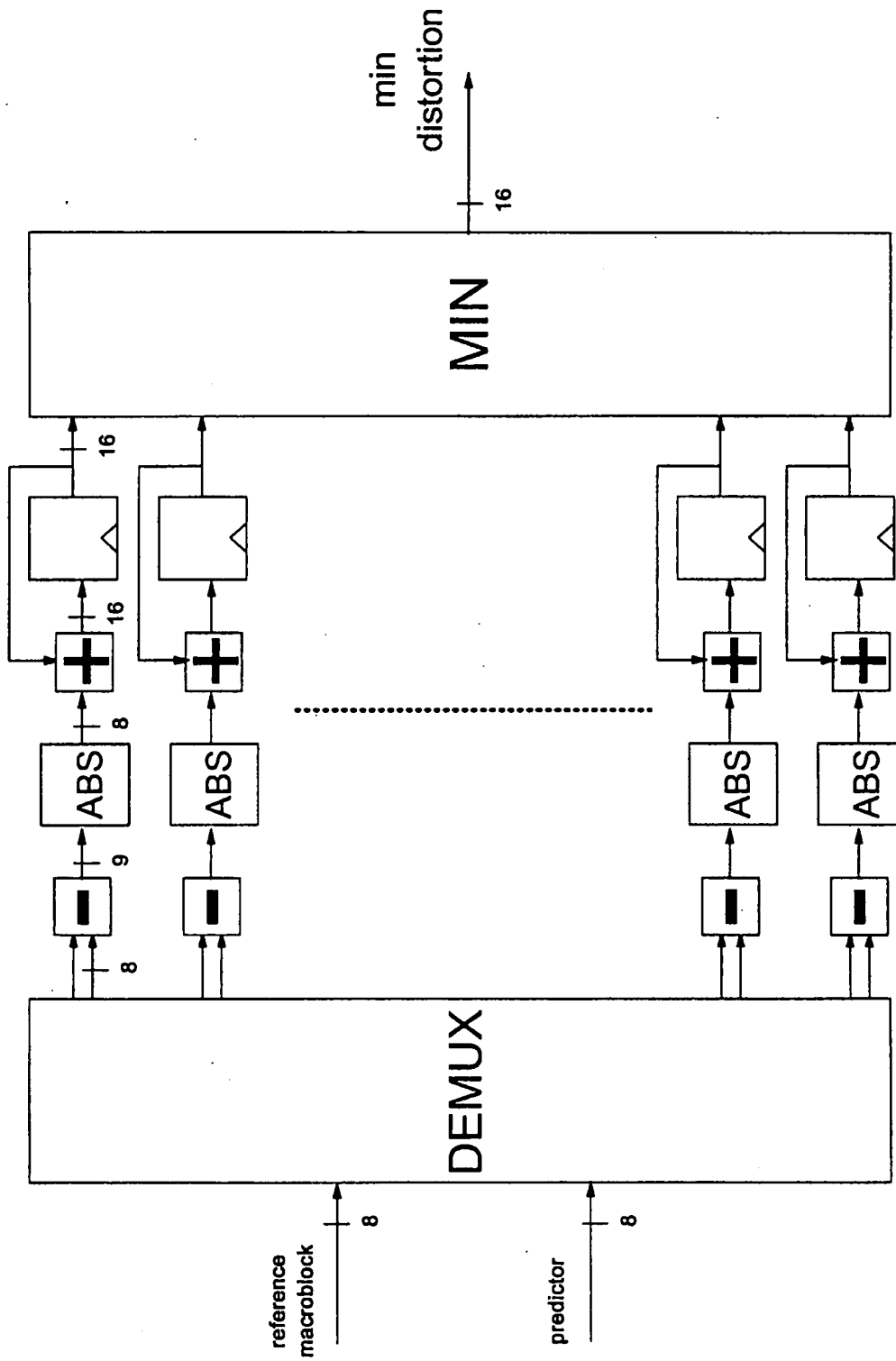


FIGURE 3

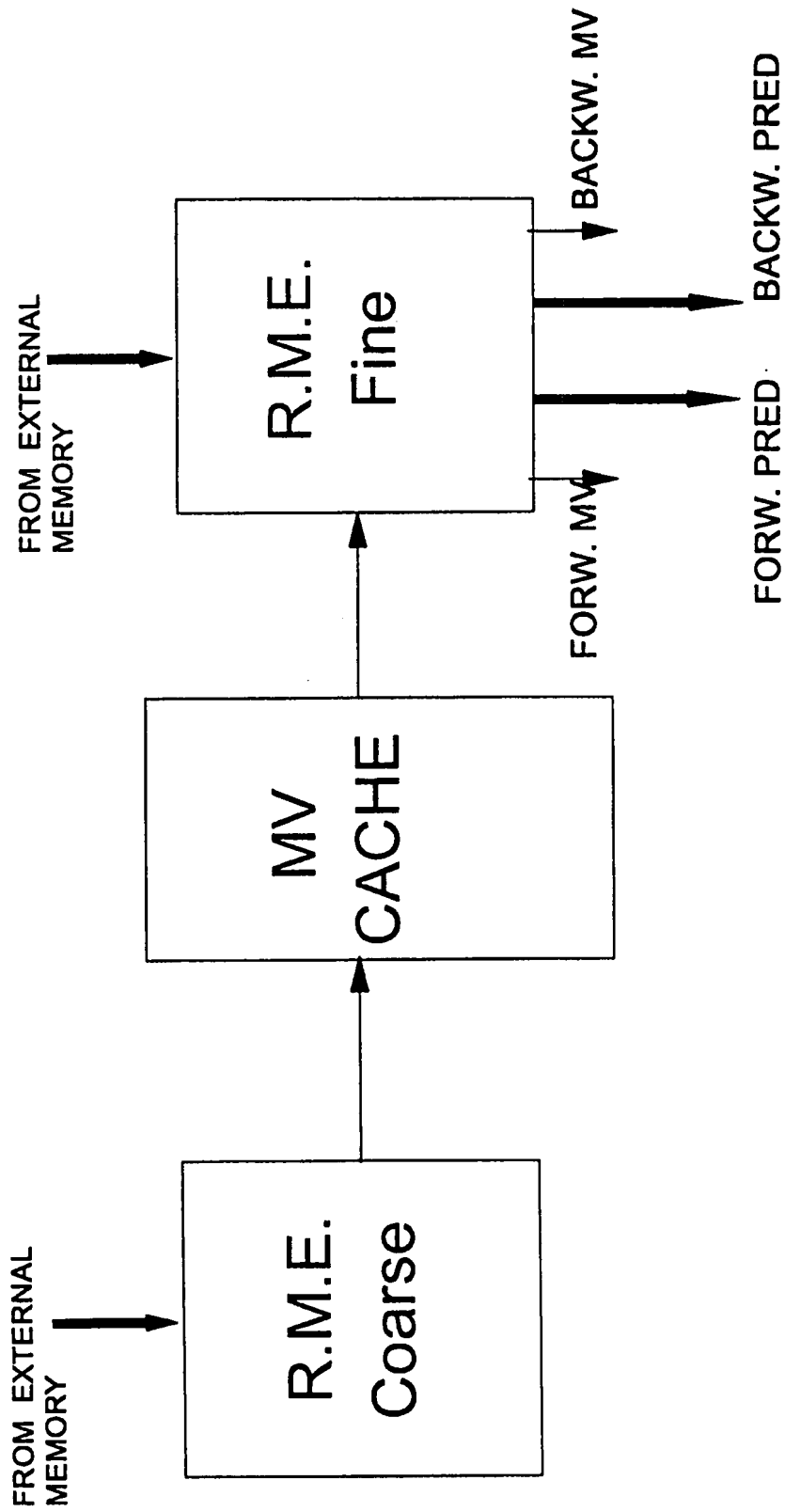


FIGURE 4

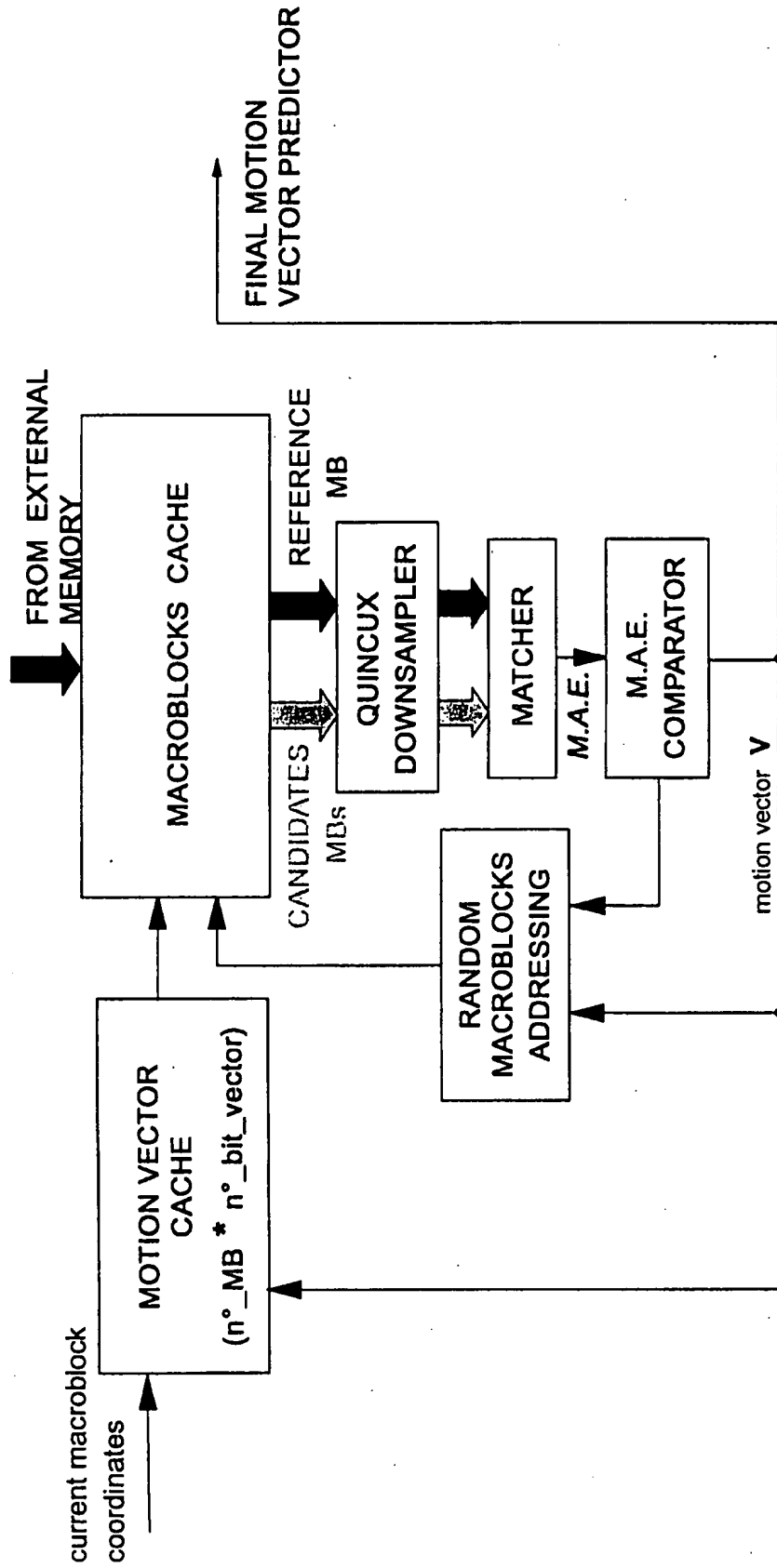


FIGURE 5

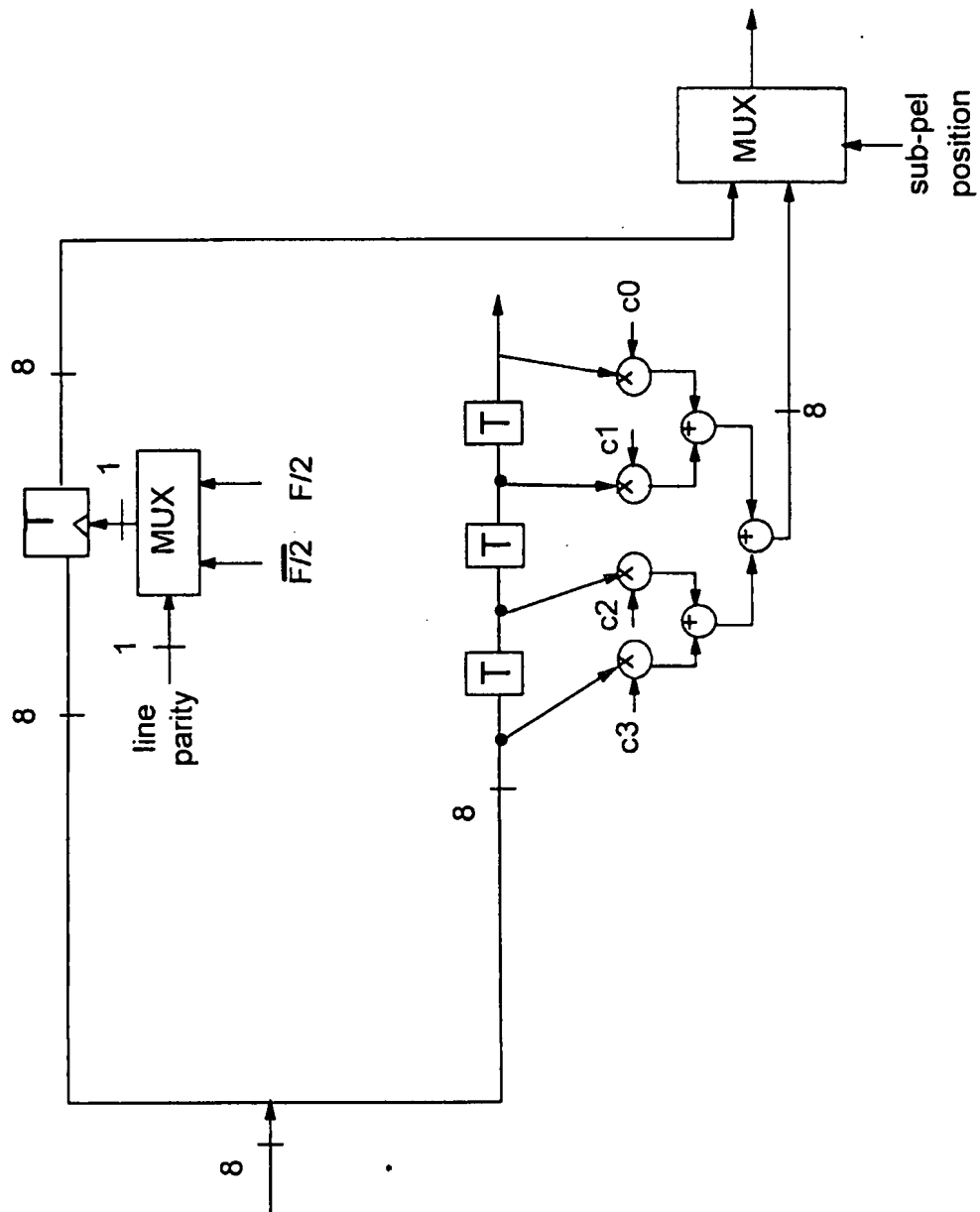


FIGURE 6

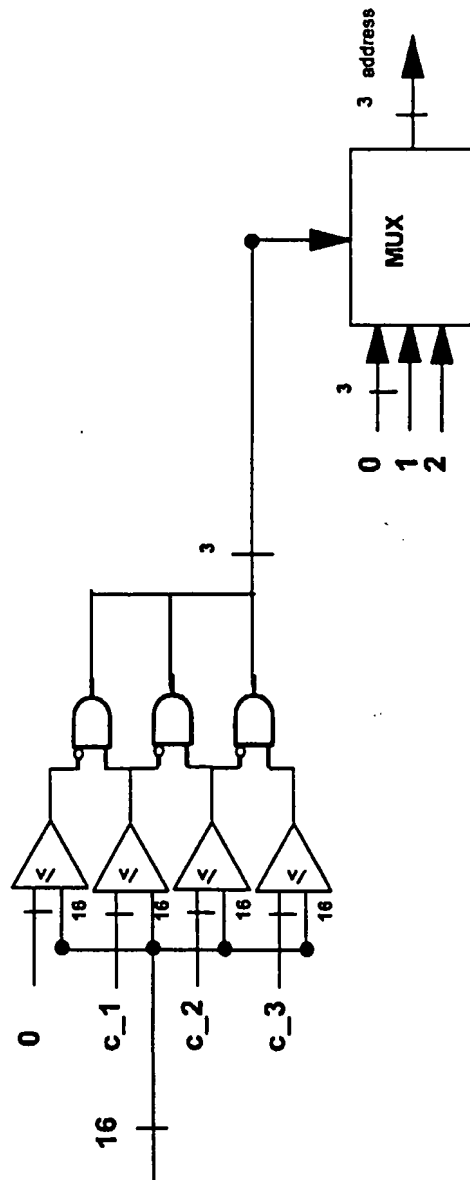


FIGURE 7

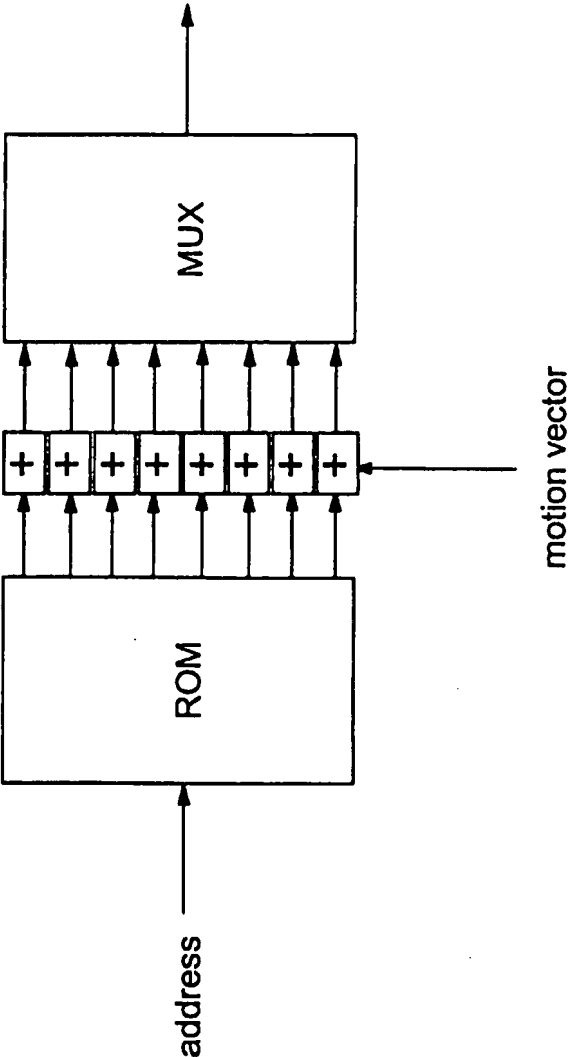


FIGURE 8

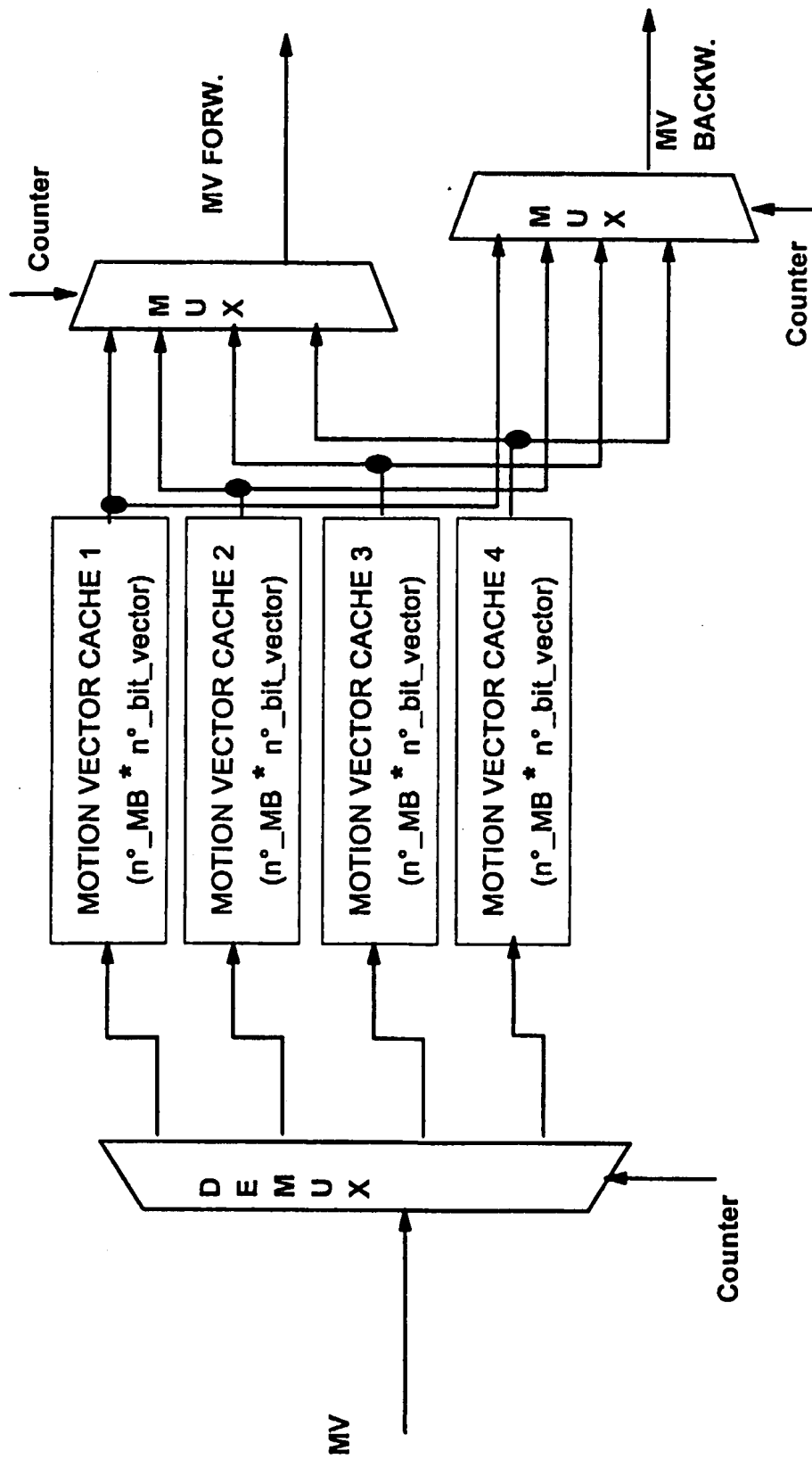


FIGURE 9

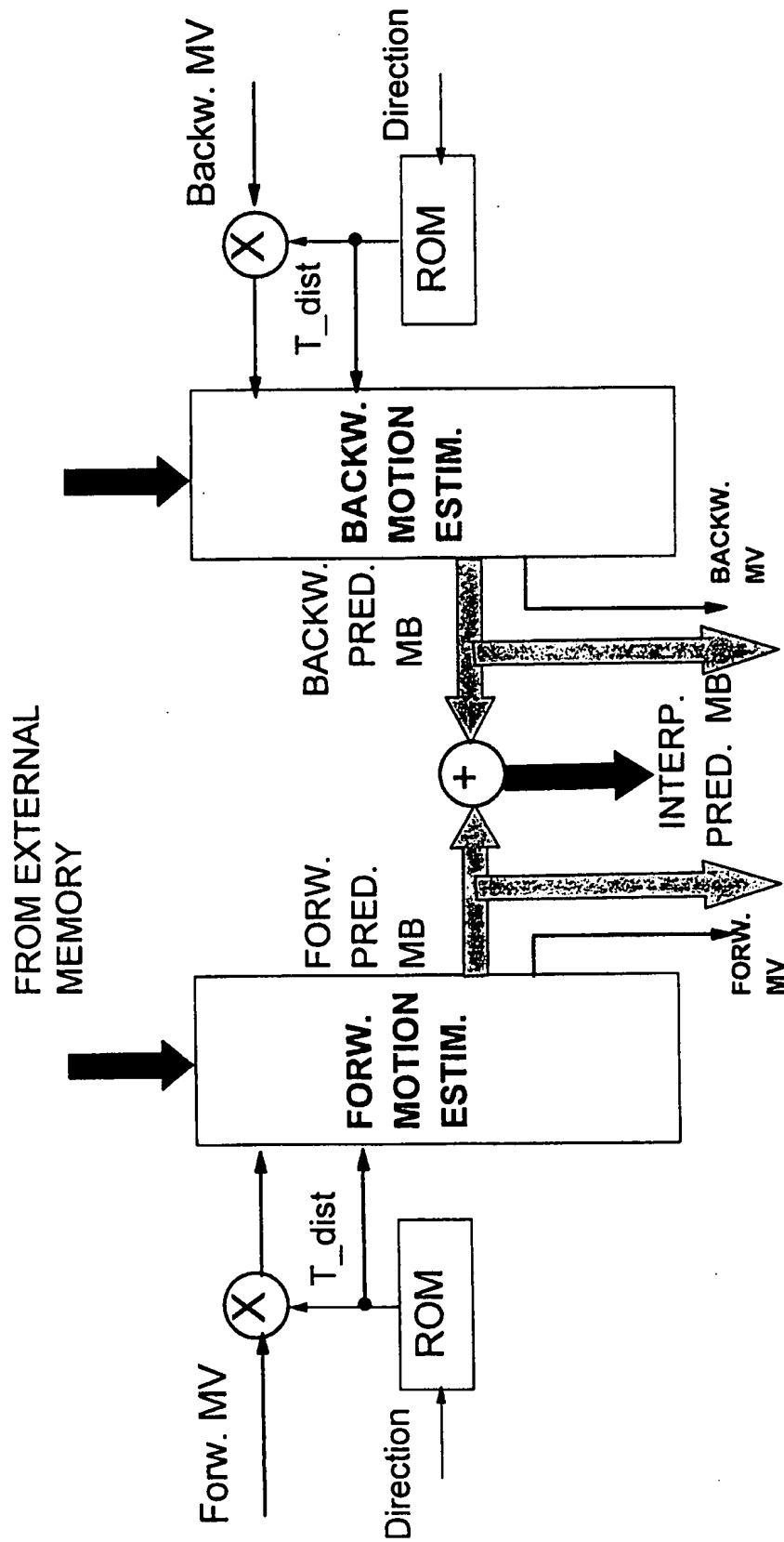


FIGURE 10

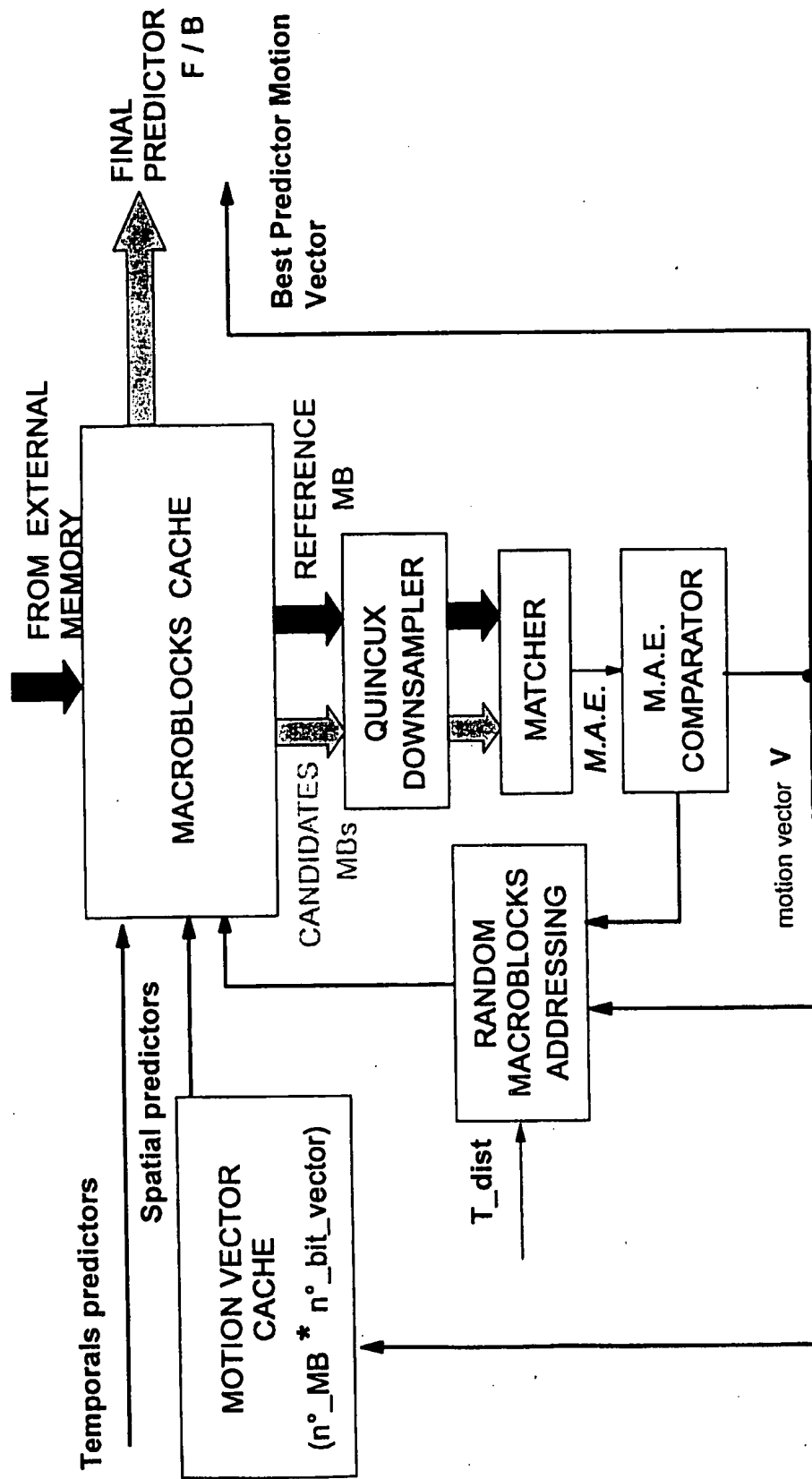


FIGURE 11

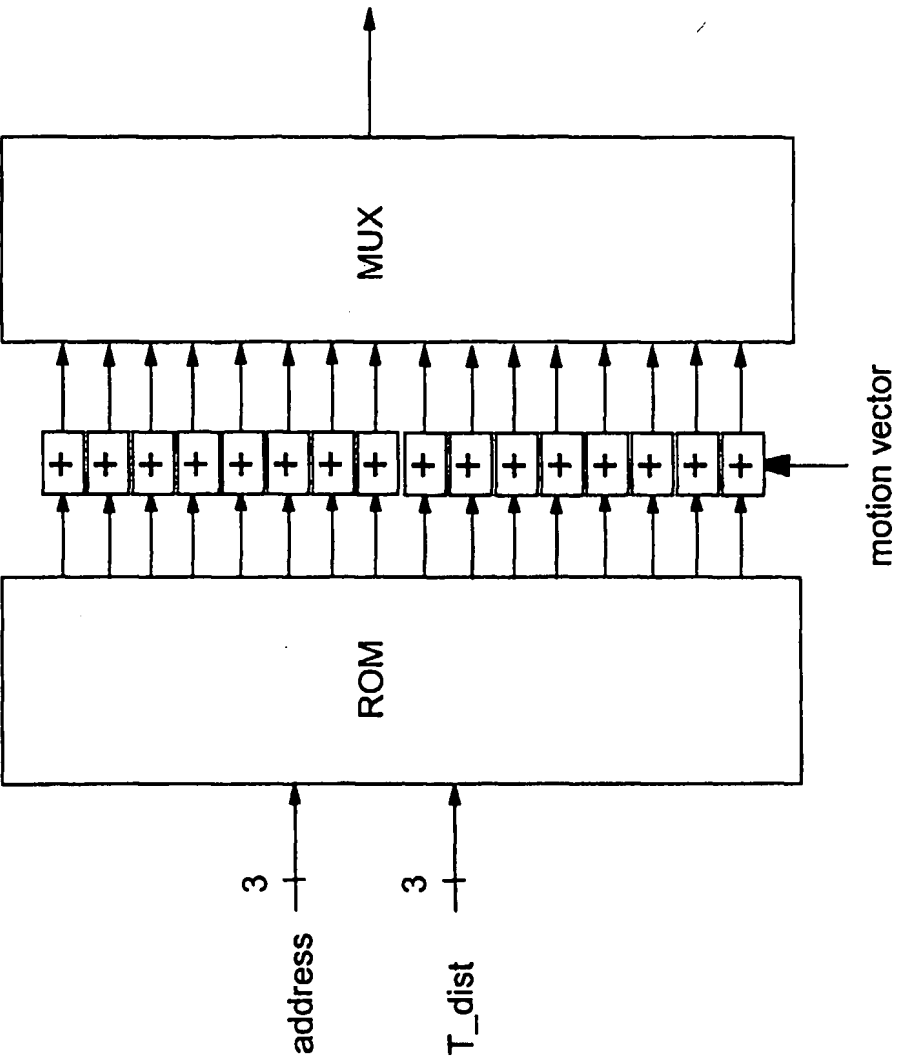


FIGURE 12



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 98 83 0163

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
A	EP 0 778 698 A (THOMSON MULTIMEDIA SA) 11 June 1997 * claims 1-3,7,8 *	1-5	H04N5/00 H04N7/36
A	US 5 049 991 A (NIIHARA TAKAMI) 17 September 1991 * abstract * * claims 1-3 *	1-5	
A	EP 0 447 068 A (VICTOR COMPANY OF JAPAN) 18 September 1991 * column 6, line 45 - column 7, line 11 *	1-5	
A	SIU-WAI WU ET AL: "JOINT ESTIMATION OF FORWARD/BACKWARD MOTION VECTORS FOR MPEG INTERPOLATIVE PREDICTION" PROCEEDINGS OF THE PICTURE CODING SYMPOSIUM (PCS), LAUSANNE, MAR. 17 - 19, 1993, no. -, 17 March 1993, pages 14.3/A-14.3/B, XP000346484 SWISS FEDERAL INSTITUTE OF TECHNOLOGY	1-5	
A	EP 0 535 746 A (PHILIPS ELECTRONICS UK LTD ; PHILIPS NV (NL)) 7 April 1993 * abstract * * column 3, line 9 - line 40 * * figure 1 *	1-5	
A	HSIEH C H ET AL: "MOTION ESTIMATION ALGORITHM USING INTERBLOCK CORRELATION" ELECTRONICS LETTERS, vol. 26, no. 5, 1 March 1990, page 276/277 XP000105589 * the whole document *	1-5	
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 7 August 1998	Examiner Berbain, F
CATEGORY OF CITED DOCUMENTS X: particularly relevant if taken alone Y: particularly relevant if combined with another document of the same category A: technological background O: non-written disclosure P: intermediate document		T: theory or principle underlying the invention E: earlier patent document, but published on, or after the filing date D: document cited in the application L: document cited for other reasons &: member of the same patent family, corresponding document	

EPO FORM 1503 (03/92) (P04001)